PSYC234: Lecture 8 post-lecture worksheet

This worksheet is to help you consolidate what you learned about the binary logistic regression during Lecture 8. It contains two activities.

This worksheet could be completed as part of the independent study hours for PSYC234. **It is optional but recommended**. **It is recommended that you complete this worksheet in advance of the WBA.**

Once you have finished, compare your answers to the answer sheet provided on Moodle. You can also use this sheet and the answer sheet for revision purposes when preparing for the class test.

Activity 1: Calculating odds ratios manually

Activity 1 will involve work with the following data. You are a researcher interested in whether being excited (yes/no) predicts whether an individual passes their driving test (yes/no). Here is a table of frequencies:

	Passed driving test	
	Yes	No
Excited - Yes	14	12
Excited - No	8	29

1. **What are the odds of passing the driving test in the "Excited – Yes" group?**

The probability of individuals who are excited (Excited – Yes) passing the driving test:

- 14/26 = 0.5384615385
- 14 is the number of participants who were excited and passed the driving test.
- 26 is the total number of individuals who responded "Excited – Yes" (14+12)

The probability of individuals who are excited (Excited – Yes) not passing the driving test:

- 12/26 = 0.4615384615
- 12 is the number of participants who were excited and did not pass the driving test.

- 26 is the total number of individuals who responded “Excited – Yes” (14+12)

The odds of individuals who are excited passing the driving test:

- 0.5384615385 / 0.4615384615 = 1.1666666668

2. **What are the odds of passing the driving test in the “Excited – No” group?**

The probability of individuals who are not excited (Excited – No) passing the driving test:

- 8/37 = 0.216216216
- 8 is the number of participants who were not excited and passed the driving test.
- 37 is the total number of individuals who responded “Excited – No” (8+29)

The probability of individuals who are excited (Excited - No) not passing the driving test:

- 29/37 = 0.783783783
- 29 is the number of participants who were not excited and did not pass the driving test.
- 37 is the total number of individuals who responded “Excited – No” (8+29)

The odds of individuals who are not excited passing the driving test:

0.216216216 / 0.783783783 = 0.275862069

3. **What is the odds ratio (where “Excited – No” is the original odds)?**

1.1666666668 / 0.275862069 = 4.23

The odds ratio = 4.23

4. **What does this odds ratio mean?**

Individuals who were excited had a 4.23x higher odds of passing the driving test relative to individuals who were not excited.

5. **Is there evidence of quasi-complete separation or complete separation here? Give a reason for your answer.**

No – all cells have quite a few observations so there is no evidence of quasi-complete separation or complete separation.

<u>Activity 2: Interpreting R output</u>

Activity 2 examines the following research question. You are a researcher interested in whether being rich (yes/no) predicts whether an individual owns a Tesla. Here is a table of frequencies:

	Has a Tesla	
	Yes	No
Rich - Yes	5	5
Rich - No	34	3

You analyse this data in R and the output of your model is below. For the outcome, you set “0” as “Has a Tesla – No” and “1” as “Has a Tesla – Yes”

```{r}
tesla_model <- glm(tesla_numeric ~ Rich, data = tesla, family=binomial())

summary(tesla_model)

```

```
Call:
glm(formula = tesla_numeric ~ Rich, family = binomial(), data = tesla)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1774  -0.4112  -0.4112  -0.4112   2.2416

Coefficients:
            Estimate Std. Error z value  Pr(>|z|)
(Intercept)  -2.4277     0.6023  -4.031 0.0000556 ***
RichYes       2.4277     0.8734   2.780   0.00544 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 42.885  on 46  degrees of freedom
Residual deviance: 34.687  on 45  degrees of freedom
AIC: 38.687

Number of Fisher Scoring iterations: 5
```

Evaluating the model:

Model statistics:

You run some code to produce the model's chi-square statistic, the degrees of freedom and the p-value. These are displayed below:

- Chi square = 8.2
- Degrees of freedom = 1
- P-value = 0.004

1. What do these values indicate?

$X^2(1) = 8.20$, $p = .004$.

This indicates that adding the “Rich” variable to our model significantly improved the fit, compared to the null model containing intercept only

Pseudo R^2:

```{r}
PseudoR2(tesla_model, which = "all")
```

```
       McFadden     McFaddenAdj        CoxSnell      Nagelkerke   AldrichNelson VeallZimmermann           Efron 
      0.1911699       0.0978972       0.1600663       0.2674640       0.1485248       0.3113013       0.2081168 
McKelveyZavoina            Tjur             AIC             BIC          logLik         logLik0              G2 
      0.2308156       0.2081168      38.6866797      42.3869749     -17.3433399     -21.4425007       8.1983218 
```

2. Which Pseudo R^2 values might you report (based on the lecture)? What is the value of these Pseudo R^2s?

McFadden, CoxSnell and Nagelkerke

- McFadden = 0.19
- CoxSnell = 0.16
- Nagelkerke = 0.27

Evaluating the individual predictors:

Looking back at the summary output:

3. What is the reference category for the predictor “Rich”?

“No” is our reference category for the Rich variable

4. What does the Intercept Estimate represent?

The log odds of someone with a Rich value of “No” having a tesla

5. What does the RichYes Estimate represent?

The change in the log odds of having a tesla value of “Yes” when going from the reference category (RichNo) to RichYes

Exponentiating the Estimates:

```{r}
tesla_model_exponentiated <- exp(tesla_model$coefficients)
tesla_model_exponentiated
```

```
(Intercept)     RichYes
 0.08823529 11.33333333
```

6. What does the Intercept represent?

The odds of having a tesla for individuals who are not rich

7. What does the RichYes value represent?

The change in odds (i.e. the odds ratio) of having a tesla (i.e. tesla = yes) when going from RichNo to RichYes

8. Interpret the RichYes value

- The odds of having a tesla are 11.33x higher if you are rich than if you are not rich

Producing confidence intervals:

```{r}
tesla_model_odds_confidence_intervals <- exp(confint(tesla_model))
tesla_model_odds_confidence_intervals
```

```
Waiting for profiling to be done...
                 2.5 %     97.5 %
(Intercept) 0.02124284  0.2452701
RichYes     2.15765202 71.8937055
```

9. What does the RichYes 95% confidence intervals represent?

- The 95% confidence around the odds ratio (for the comparison between RichNo to RichYes).

10. From the p-value in the summary table for the RichYes row, what can you conclude?

- Whether an individual is rich (yes/no) significantly predicts whether they have a tesla (yes/no; p = .005)